Homework III
Theoretical Statistics/Machine Learning

Sagar Ghosh
Department of Statistics and Data Sciences
The University of Texas at Austin
Email: sg63684@my.utexas.edu

February 24, 2026

1 More Sub-Gaussian Charecterizations.

X is a real-valued random variable and we define its ψ_2(orlicz) norm as

$$\|X\|_{\psi_2} = \inf\{K > 0 : \mathbb{E}\left(\exp(X^2/K^2)\right) \leq 2\}.$$

X is sub-Gaussian iff $\|X\|_{\psi_2} < \infty$.

(a) **Sub-Gaussianity via (almost) Gaussian tail domination.** Let $g \sim \mathcal{N}(0,1)$. For two random variables U, V define the relation

$$U \preccurlyeq V \iff \mathbb{P}[|U| \geq t] \leq 2\mathbb{P}[|V| \geq t] \quad \text{for all } t \geq 0.$$

We will show that X is sub-Gaussian iff $\exists\, K > 0$ such that $X \preccurlyeq Kg$.

(i) Suppose $\|X\|_{\psi_2} \leq L \implies \mathbb{E}[\exp(X^2/L^2)] \leq 2$. Therefore using Markov's inequality, we get $\mathbb{P}[|X| \geq t] = \mathbb{P}[\exp(X^2/L^2) \geq \exp(t^2/L^2)] \leq 2\exp(-t^2/L^2)$. Now to compare with the Gaussian tails, we use the Gordon's bound on Mill's ratio for the standard gaussian tail: $\mathbb{P}[g \geq u] \leq \frac{\phi(u)}{u+u^{-1}}$, where $\phi(u) = \frac{1}{\sqrt{2\pi}}\exp(-u^2/2)$. Hence by symmetry, $\mathbb{P}[|g| \geq u] = 2\mathbb{P}[g \geq u] \geq 2\frac{\phi(u)}{u+u^{-1}}$. Next, we substitute $u = t/(CL)$ to see that

$$\mathbb{P}[|g| \geq t/(CL)] \geq 2\frac{1}{\frac{t}{CL} + \frac{CL}{t}} \exp\left(-\frac{1}{2}\frac{t^2}{C^2L^2}\right) = 2\exp\left[-\log\left(\frac{t}{CL} + \frac{CL}{t}\right) - \left(\frac{1}{2}\frac{t^2}{C^2L^2}\right)\right].$$

So, now it only remains to bound $-\log\left(\frac{t}{CL} + \frac{CL}{t}\right)$. Let $x = \frac{t}{CL} + \frac{CL}{t}$. Bu abuse of notation, consider $f(x) = x^2 - \log(x + 1/x)$. Then $f'(x) = 2x - \frac{x}{x^2+1} = x\left[2 - \frac{1}{x^2+1}\right] \geq 0$ for $x > 0$. So, for $x \geq 1$, $f(x) \geq f(1) = 1 - \log 2 > 0 \implies \log(x + 1/x) > -x^2$. Hence for $t/(CL) \geq 1$, we get

$$\mathbb{P}[|g| \geq (t/CL)] \geq 2\exp\left(-\frac{t^2}{C^2L^2} - \frac{1}{2}\frac{t^2}{C^2L^2}\right) = 2\exp\left(-c\frac{t^2}{L^2}\right) \geq \mathbb{P}[|X| \geq t].$$

Note that for $x = t/(CL) \leq 1$, we can use some numerical bounds (up to a constant). Since we are interested in tail probabilities, we can assume wlog that $\exists\delta > 0$ such that $t > \delta$, otherwise we would estimate $\mathbb{P}[|X| \geq 0] = 1$, is of no interest. Therefore for $\delta/(CL) \leq x = t/(CL) \leq 1$, we have $\mathbb{P}[|g| \geq (t/CL)] \geq 2\left[\frac{1}{2} - \frac{1}{2(\delta^2+1)}\right]\exp(-t^2/(C^2L^2)) \geq 2\exp(-c't^2/L^2) \geq \mathbb{P}[|X| \geq t]$,

where the first inequality follows because $\frac{x}{x^2+1} \geq \left[\frac{1}{2} - \frac{1}{2(\delta^2+1)}\right]$ and c' is a constant depending on δ.

(ii) If $X \preccurlyeq Kg$, then $\mathbb{P}[|X| \geq t] \leq 2\mathbb{P}[|g| \geq t/K] \leq 4\exp(-t^2/(2K^2))$, where we used that for a standard

Gaussian variable Z, $\mathbb{P}[|Z| \geq u] \leq 2\exp(-u^2/2)$. Therefore, using the tail integral definition of expectation,

$$\begin{aligned}
\mathbb{E}[\exp(X^2/(CK^2))] &= \int_0^\infty \mathbb{P}[\exp(X^2/CK^2) \geq t]dt \\
&= \int_0^\infty \mathbb{P}[|X| \geq CK \log t]dt \\
&\leq 4\int_0^\infty \exp\left[-\frac{\sqrt{C}(\log t)^2}{2}\right] dt \\
&= 4\int_{-\infty}^\infty \exp(-cz^2)\exp(z)dz \quad \text{[by substituting } t = e^z\text{]} \\
&\lesssim 2 \quad \text{[Since the integrand is finite, involves some Gamma function]},
\end{aligned}$$

completing the proof that $\|X\|_{\psi_2} \lesssim K$.

(b) Now we will remove the factor 2 from the definition of $\preccurlyeq$ and see that (a) can fail. To see this, set $X = \epsilon$, where $\epsilon \sim$ Rademacher$\{\pm 1\}$. Choose $t = 1$. Then $\mathbb{P}[|X| \geq 1] = \mathbb{P}[|\epsilon| \geq 1] = 1$. On the contrary, $\mathbb{P}[|Kg| \geq 1] = \mathbb{P}[|g| \geq 1/K] < 1$ for $\infty > K > 0$. Hence, $\mathbb{P}[|X| \geq 1] \nleq \mathbb{P}[|Kg| \geq 1]$. If we have the 2 factor present, then we need to choose K such that $2\mathbb{P}[|Kg| \geq 1] \geq 1$ or $\mathbb{P}[|g| \geq K] \geq 1/2$. For discrete numerical choice for K, the following simulation in R reveals the choice for K:

```
> 1/qnorm(0.6)
[1] 3.947154

```

This means, for $K \approx 3.95$, $\mathbb{P}[|g| \geq 1/3.95] \approx \mathbb{P}[|g| \geq 0.255] \geq 2(1-0.6) = 0.8 > 1/2$. Therefore, $K \approx 3.95$ may be the choice for which we can say that ϵ is sub-Gaussian with the 2 factor present.

(c) **A local MGF of X^2 Charecterizations.**

(i) Suppose $\|X\|_{\psi_2} \leq L \implies \mathbb{E}[\exp(X^2/L^2)] \leq 2$. Now using Jensen's inequality to this concave function $x \to x^\theta$ for $0 \leq \theta \leq 1$, we get $\mathbb{E}[Y^\theta] \leq [\mathbb{E}Y]^\theta$, putting $Y = \exp(X^2/L^2)$ and $\theta = \lambda^2 L^2$, we get for $|\lambda| \leq 1/L$, $\mathbb{E}[\exp(\lambda^2 X^2)] = \mathbb{E}[\exp(X^2/L^2)^{\lambda^2 L^2}] \leq \mathbb{E}[\exp(X^2/L^2)]^{\lambda^2 L^2} \leq 2^{\lambda^2 L^2} = \exp(\log 2\lambda^2 L^2) \leq \exp(\lambda^2 K^2)$, we can choose $K \approx \log 2L$ for this specific choice.

(ii) Assuming that the inequality $\mathbb{E}[\exp(\lambda^2 X^2)] \leq \exp(\lambda^2 K^2)$ holds for $|\lambda| \leq 1/K$, choosing $\lambda = 1/(2K)$, we get $\mathbb{E}[\exp(X^2/(4K^2))] \leq \exp(K^2/(4K^2)) = \exp(1/4) < 2$, since $4\log_e 2 = \log_e(2^4) > 1 \implies \log_e 2 > 1/4 \implies 2 > \exp(1/4)$. Hence, by the definition of ψ_2 Orlicz norm, we have $\|X\|_{\psi_2} \leq 2K$.

2 Sub-Exponential Charecterizations.

Z is a real valued random variable with $\mathbb{E}Z = 0$, We define the centred $\log -$ MGF as $\psi_Z(\lambda) = \log \mathbb{E}[e^{\lambda Z}] \in (-\infty, \infty), \lambda \in \mathbb{R}$. We also define the ψ_1 (Orlicz) Norm as $\|Z\|_{\psi_1} = \inf\{K > 0 \mid \mathbb{E}\exp(|Z|/K) \leq 2\}$.

(a) We will prove the equivalence of four sub-exponential properties:

(a_1) **(i) $\implies$ (ii): Tail bound implies moment growth.** Fix $K > 0$, we assume for all $t \geq 0$, $\mathbb{P}[|Z| \geq t] \leq 2\exp(-t/K)$. We will use the formulation: $\mathbb{E}|X|^p = p\int_0^\infty t^{p-1}\mathbb{P}[|X| \geq t]dt$. Hence,

$$\begin{aligned}
\mathbb{E}[|Z|^p] &= p\int_0^\infty t^{p-1}\mathbb{P}[|Z| \geq t]dt \\
&\leq 2p\int_0^\infty t^{p-1}\exp(-t/K)dt \\
&= 2pK^p\int_0^\infty z^{p-1}\exp(-z)dz \quad \text{[substituting } t = Kz\text{]} \\
&= 2pK^p\Gamma(p) \\
&= 2K^p\Gamma(p+1) \\
&\lesssim e^pK^pp^p,
\end{aligned}$$

Since $\Gamma(p+1) \lesssim (p)^p$. Taking the $p-$ th root on both sides, we get $\mathbb{E}[|Z|^p]^{1/p} \leq cKp$, where c is a constant and $c < 1/e$, independent of K and p.

(a_2) **(ii) $\implies$ (iii): Moment growth implies finite Orlicz norm.** Using the series expnsion of $\exp z$, we write

$$\begin{aligned}
\mathbb{E}\exp(|Z|/K) &= \sum_{r=0}^\infty \mathbb{E}\frac{|Z|^r}{K^r r!} \\
&\leq \sum_{r=0}^\infty \frac{C^rK^rr^r}{K^r r!} \\
&\leq \sum_{r=0}^\infty (Ce)^r \quad \text{[using Stirling bound: } p! \geq (p/e)^p\text{]} \\
&\leq c' \quad \text{[Since } C < 1/e\text{]} \\
&\lesssim 2,
\end{aligned}$$

i.e., The Orlicz norm of Z is finite.

(a_3) **(iii) $\implies$ (iv): Finite Orlicz Norm implies Local Quadratic Log-MGF.** We know that for small $c > 0$, when $|\lambda| < c/K$, $\exp(\lambda Z) \leq 1 + \lambda Z + \lambda^2 Z^2$. Also, finite Orlicz norm implies $\mathbb{E}(|Z|^2/K^2) \leq \mathbb{E}[\exp(|Z|/K)] \leq 2 \implies \mathbb{E}[Z^2] \leq 2K^2$. Therefore, using the local boundedness of exponential function by a quadratic function, we get $\mathbb{E}[e^{\lambda Z}] \leq \mathbb{E}[1 + \lambda Z + \lambda^2 Z^2] \leq 1 + \lambda^2 2K^2 \leq \exp(C^2\lambda^2 K^2)$. Taking the log on either sides, we get

$$\psi_Z(\lambda) = \log \mathbb{E}[e^{\lambda Z}] \leq \log(1 + C\lambda^2 K^2) \leq C\lambda^2 K^2.$$

(a_4) **(iv) $\implies$ (i): Local quadratic log MGF implies exponential tail bound:** Using Chernoff bound, we get $\mathbb{P}[Z \geq t] \leq \inf_\lambda \exp(-\lambda t + \psi_Z(\lambda)) \leq \inf_\lambda \exp(-\lambda t + C\lambda^2 K^2)$ [The last inequality follows from the local quadratic log-MGF bound for $|\lambda| \leq c/K$]. Therefore, we optimize $-\lambda t + CK^2\lambda^2$ on $[-c/K, c/K]$. The unconstrained optimal value of $-\lambda t + C\lambda^2 K^2$ is $-t^2/(2K^2 C)$, and the extreme value of the constrained optimization is $-ct/K + Cc^2$, and for a tail bound this has to be true for all $t \geq 0$. Since for large t, the tail bound with smaller exponent in t will dominate [since it is in the argument of an exponential], we get that for $|\lambda| \leq c/K$, the optimal value would be $\exp(-c't/K)$, for some constant c'. Considering the other tail probability $\mathbb{P}[Z \leq -t]$, and combining them together we get

$$\mathbb{P}[|Z| \geq t] \leq 2\exp(-c't/K) \lesssim 2\exp(-t/K).$$

(b) **Comparison with Sub-Gaussianity.**

Suppose Z is σ sub-Gaussian, this means the moment bound for Z is $\mathbb{E}[|Z|^p]^{1/p} \leq C\sigma\sqrt{p} \leq C\sigma p$, since $p \geq 1$. Therefore, suing a similar argument to $(ii) \implies (iii)$ in part (a), we can write $\mathbb{E}[\exp(|Z|/\sigma)] \leq 2$. This means Z has a finite ψ_1 norm, which is in the order of σ, i.e., $\|Z\|_{\psi_1} \leq C\sigma$. Hence Z is σ sub-exponential.

For the other direction, consider $X \sim \text{Exp}(1)$. Center the variable $Z = X - \mathbb{E}X = X - 1$. Therefore for any $t \geq 1$ [since we are interested in tail probability, it's enough to show that the sub-Gussianity of Z fails for sufficiently large t],

$$\mathbb{P}[|Z| \geq t] = \int_t^\infty \mathbb{P}[|Z| = r]dr \geq \int_{t+1}^\infty \mathbb{P}[X = r]dr = \int_{t+1}^\infty e^{-r}dr = e^{-(t+1)},$$

Where the second equality follows from the fact that $t \geq 1$. We note that this tail probability is $\mathcal{O}(e^{-t})$, this is strictly larger than the order e^{-t^2}. And Z is sub-exponential, since

$$\mathbb{P}[|Z| \geq t] = \int_t^\infty \mathbb{P}[|Z| = r]dr \leq 2\int_{t+1}^\infty \mathbb{P}[X = r]dr = 2\int_{t+1}^\infty e^{-r}dr = 2e^{-(t+1)} \lesssim ce^{-t}.$$

3 Bounding ℓ_p norms of Random Vectors.

Let $X_2, ..., X_n$ be independent random variables with $\mathbb{E}[X_i] = 0$. We assume they are sub-Gaussian and set $K = \max_{1\le i\le n} \|X_i\|_{\psi_2}$. Let $X = (X_1, ..., X_n) \in \mathbb{R}^n$ and set $p \in [1, \infty)$. Define $\|X\|_p = (\sum_{i=1}^n |X_i|^p)^{1/p}$ and $\|X\|_\infty = \sup_{1\le i\le n} |X_i|$. Throughout we will assume sub-Gaussian moment growth and tail bounds: $\mathbb{E}[|X_i|^q]^{1/q} \le CK\sqrt{q}$ and $\mathbb{P}[|X_i| \ge t] \le \exp(-ct^2/K^2)$.

(a) **Expected ℓ_p norms for finite p.** Using the concavity of the function $x \to x^{1/p}$ and Jensen's inequality we get [we set $u = \|X\|_p^p$],

$$\mathbb{E}[\|X\|_p] = \mathbb{E}\left[\sum_i |X_i|^p\right]^{1/p} \le \left[\sum_i \mathbb{E}[|X_i|^p]\right]^{1/p} \le \left[\sum_i C^p K^p p^{p/2}\right]^{1/p} = ((CK\sqrt{p})^p n)^{1/p} = CKn^{1/p}\sqrt{p}.$$

(b) **Expected ℓ_∞ norm.** Note that using union bound and the sub-Gaussian tail bound, we can write

$$\mathbb{P}[\|X\|_\infty \ge r] = \mathbb{P}[\max_i |X_i| \ge r] \le \mathbb{P}[\cup_i\{|X_i| \ge r\}] \le \sum_i \mathbb{P}[|X_i| \ge r] \le 2n\exp(-cr^2/K^2) = 2\exp[\log n - cr^2/K^2].$$

Now we will choose r such that the right side becomes in the form of $\exp(-ct^2)$. We choose $r = CK(\sqrt{\log n} + t) \implies r^2 \ge C^2K^2\log n + C^2K^2t^2 \implies -\frac{1}{C^2}r^2/K^2 + \log n \le -t^2$. Hence, $\mathbb{P}[\|X\|_\infty \ge CK(\sqrt{\log n} + t)] \le 2\exp(-ct^2)$, where we can take $c = 1$, proving the claim. For the expectation, we will use tail integral $\mathbb{E}[Y] = \int_0^\infty \mathbb{P}[Y \ge r]dr$ for $Y \ge 0$. Then

$$\begin{aligned}
\mathbb{E}[\|X\|_\infty] &= \int_0^\infty \mathbb{P}[\|X\|_\infty \ge r]dr \\
&= \int_{-\sqrt{\log n}}^\infty \mathbb{P}[\|X\|_\infty \ge CK(\sqrt{\log n} + t)]CKdt \quad [\text{substituting } r = CK(\sqrt{\log n} + t)] \\
&= CK\int_{-\sqrt{\log n}}^\infty \exp(-t^2)dt \quad [c = 1] \\
&= CK\int_{-\sqrt{\log n}}^0 \exp(-t^2)dt + CK\int_0^\infty \exp(-t^2)dt \\
&\le CK\sqrt{\log n} + C'K \quad [\text{Since } \exp(-t^2) \le 1 \text{ and the second integral is a constant}] \\
&\le C_1K\sqrt{\log n},
\end{aligned}$$

where C_1 is some constant depending on C, C'.

(c) **Two Regimes of Expected ℓ_p norms.** We will use the equivalence of ℓ_p norms in finite (but arbitrary) dimensions: $\|X\|_\infty \le \|X\|_p \le n^{1/p}\|X\|_\infty$. For $p \le \log n$, we have $\mathbb{E}[\|X\|_p] \le CK\sqrt{p}n^{1/p}$ from part (a). And for $p \ge \log n$, we have $\frac{1}{p} \le \frac{1}{\log n} \implies n^{1/p} \le n^{1/\log n} = e$. Therefore, for $p \ge n$, using $\|X\|_p \le n^{1/p}\|X\|_\infty$, using part (b), we write

$$\mathbb{E}[\|X\|_p] \le n^{1/p}\mathbb{E}[\|X\|_\infty] \le n^{1/p} CK\sqrt{\log n} \le eCK\sqrt{\log n},$$

as desired.

(d) **High-probability bound for ℓ_p norms when $p \ge \log n$.** Since for $p \ge \log n$, $\|X\|_p \le n^{1/p}\|X\|_\infty \le e\|X\|_\infty$, we have

$$\mathbb{P}[\|X\|_p \ge CK(\sqrt{\log n} + t)] \le \mathbb{P}[\|X\|_\infty \ge \frac{1}{e}CK(\sqrt{\log n} + t)]2 \le \exp(-ct^2/e^2) = 2\exp[-c't^2].$$

(e) **High-probability bound for ℓ_p norms when $p \le \log n$.** Fix $p \in [1, \log n]$. We write $\|X\|_p^q = (\sum_i |X_i|^p)^{q/p} \le n^{q/p-1}\sum_i |X_i|^q$. Taking expectation on wither side and using $\mathbb{E}[|X_i|^q] \le C^q K^q q^{q/2}$, we get

$$\left(\mathbb{E}[\|X\|_p^q]\right)^{1/q} \le (n^{q/p} C^q K^q q^{q/2})^{1/q} = CKn^{1/p}\sqrt{q}.$$

(f) **Thin Cell for $\|X\|_2$ (sharper than part (d)).**

Assume $X_1, \ldots, X_n$ are independent, mean-zero, sub-Gaussian random variables with $\mathbb{E}X_i^2 = 1, \quad \|X_i\|_{\psi_2} \le K$. Let $X = (X_1, \ldots, X_n)$. We know that for X_i s being sub-Gaussian, $Z_i := X_i^2 - 1$ are centered sub-exponential random variables satisfying $\|Z_i\|_{\psi_1} \le CK^2$. Additionally, $\|X\|_2^2 - n = \sum_{i=1}^n Z_i$. Applying Bernstein's inequality for sums of independent sub-exponentials gives, for all $u \ge 0$,

$$\mathbb{P}\left(\left|\frac{\|X\|_2^2}{n} - 1\right| \ge u\right) \le 2\exp\left(-cn\min\left(\frac{u^2}{K^4}, \frac{u}{K^2}\right)\right).$$

Using the elementary implication $\left|\frac{\|X\|_2}{\sqrt{n}} - 1\right| \ge \delta \;\Rightarrow\; \left|\frac{\|X\|_2^2}{n} - 1\right| \ge \max\{\delta, \delta^2\}$, we set $\delta = t/\sqrt{n}$. Then

$$\mathbb{P}\big(|\|X\|_2 - \sqrt{n}| \ge t\big) \le 2\exp\left(-cn\min\left(\frac{\max(\delta,\delta^2)^2}{K^4}, \frac{\max(\delta,\delta^2)}{K^2}\right)\right).$$

A direct check of the cases $\delta \le 1$ and $\delta \ge 1$ gives

$$n\min\left(\frac{\max(\delta,\delta^2)^2}{K^4}, \frac{\max(\delta,\delta^2)}{K^2}\right) \ge c'\frac{t^2}{K^4}.$$

Therefore, combining all these we get

$$\mathbb{P}\big(|\|X\|_2 - \sqrt{n}| \ge t\big) \le 2\exp\left(-c\frac{t^2}{K^4}\right).$$

4 Bounding L_p norms of linear forms (Khintchine inequalities).

Throughout we assume $X_1, ..., X_N$ are independent random variables and $a = (a_1, ..., a_N) \in \mathbb{R}^N$. We write

$$S(a) = \sum_{i=1}^{N} a_i X_i, \quad \|a\|_2 = \left(\sum_{i=1}^{N} a_i^2\right)^{1/2}, \quad \|a\|_\infty = \max_{1\leq i\leq N} |a_i|.$$

We also assume $C, c > 0$ are absolute constants.

(a) **Sub-Gaussian Khintchine L_p Bound ($p \geq 2$).** We assume $\mathbb{E}[X_i] = 0, \mathbb{E}[X_i^2] = 1, \|X_i\|_{\psi_2} \leq K_2$ for all i, and for some $K_2 \geq 1$. We know that for any random variable Y, $\|Y\|_{L_p} \geq \|Y\|_{L_q}$ for $p \geq q \geq 1$. Hence,

$$\|S(a)\|_{L_p} \geq \|S(a)\|_{L_2} = \left[\mathbb{E}\left(\sum_{i=1}^{N} a_i X_i\right)^2\right]^{1/2} = \left[\sum_{i=1}^{N} a_i^2 \mathbb{E}[X_i^2]\right]^{1/2} = \left[\sum_{i=1}^{N} a_i^2\right]^{1/2} = \|a\|_2,$$

where the second equality follows from independence. Since X_is are sub-gaussian, $a_i X_i$s are also sub-Gaussian with $\|a_i X_i\|_{\psi_2} \leq |a_i| K_2$. Now a sub-Gaussian random variable Y always satisfies $\mathbb{E}[e^{\lambda Y}] \leq \exp(c\lambda^2 K^2)$ where $\|Y\|_{\psi_2} \leq K$ [This is also an if-and-only-if like condition]. For a sum of independent random variables like $S(a)$, we have $\mathbb{E}[e^{\lambda S(a)}] = \prod_{i=1}^{N} \mathbb{E}[e^{\lambda a_i X_i}] \leq \exp\left(\sum_{i=1}^{N} c\lambda^2 a_i^2 K_2^2\right) = \exp\left(c\lambda^2 K_2^2 \|a\|_2^2\right)$. Hence, $\|S(a)\|_{\psi_2} \leq CK_2\|a\|_2$. Finally we apply the standard implication $\|Y\|_{L_p} \leq c\sqrt{p}\|Y\|_{\psi}$. Hence, $\|S(a)\|_{L_p} \leq c\sqrt{p}\|S(a)\|_{\psi_2} \leq cC\sqrt{p}K_2\|a\|_2 = c' K_2 \sqrt{p}\|a\|_2$.

(b) **Sub-Exponential Bernstein-Khintchine Tail Bound.** This is a direct application of Bernstein type of inequalities for sub-exponential sums. For $Y_1, Y_2, ..., Y_N$ independent mean-zero sub-exponential random variables, then

$$\mathbb{P}\left[\left|\sum_{i=1}^{N} Y_i\right| \geq t\right] \leq 2\exp\left[-c\min\left\{\frac{t^2}{V}, \frac{t}{B}\right\}\right],$$

where $V = \sum_{i=1}^{N} \|Y_i\|_{\psi_1}^2$ and $B = \max_{1\leq i\leq N} \|Y_i\|_{\psi_1}$. We apply this result with $Y_i = a_i X_i$, then $S(a) = \sum_i Y_i$. $V = \sum_i |a_i X_i|_{\psi_1}^2 \leq K_1^2 \|a\|_2^2$, and $B = \max_{1\leq i\leq n} |a_i X_i|_{\psi_1} \leq \max_{1\leq i\leq n} |a_i||X_i|_{\psi_1} \leq K_1 \max_{1\leq i\leq n} |a_i| = K_1\|a\|_\infty$. With all these in hand, we get the required Bernstein-Khintchine inequality

$$\mathbb{P}\left[|S(a)| \geq t\right] \leq 2\exp\left[-c\min\left\{\frac{t^2}{V}, \frac{t}{B}\right\}\right] \leq 2\exp\left[-c\min\left\{\frac{t^2}{K_1^2\|a\|_2^2}, \frac{t}{K_1\|a\|_\infty}\right\}\right],$$

Since $V \leq K_2^2\|a\|_2^2$ and $B \leq K_1\|a\|_\infty$, $-c\min\left\{\frac{t^2}{V}, \frac{t}{B}\right\} \leq -c\min\left\{\frac{t^2}{K_1^2\|a\|_2^2}, \frac{t}{K_1\|a\|_\infty}\right\}$.

(c) **Sub-Exponential Bernstein-Khintchine L_p bound.** We will use the expression for the expectation of a random variable Y, which is $\mathbb{E}[|Y|^p] = p\int_0^\infty t^{p-1}\mathbb{P}[|Y| \geq t]dt$. Putting $Y = S(a)$ and using the tail bound from part (b) and the fact $e^{-\min(u,v)} \leq e^{-u} + e^{-v}, u, v \geq 0$, we get

$$\begin{aligned}
\|S(a)\|_{L_p}^p = \mathbb{E}[|S(a)|^p] &\leq p \int_0^\infty t^{p-1} 2 \exp\left[-c \min\left\{\frac{t^2}{K_1^2\|a\|_2^2}, \frac{t}{K_1\|a\|_\infty}\right\}\right] dt \\
&\leq 2p \int_0^\infty t^{p-1} \exp\left[-c\frac{t^2}{K_1^2\|a\|_2^2}\right] dt + 2p \int_0^\infty t^{p-1} \exp\left[-c\frac{t}{K_1\|a\|_\infty}\right] dt \\
&= 2p \left(\frac{K_1\|a\|_2}{\sqrt{c}}\right)^p \Gamma((p+1)/2) + 2p \left(\frac{K_1\|a\|_\infty}{c}\right)^p \Gamma(p) \\
&\leq c[eK_1\|a\|_2]^p p^{p/2} + c[eK_1\|a\|_\infty]^p p^p,
\end{aligned}$$

for some constant $c > 0$, where we also used that $p \leq e^p$ for all $p \geq 2$. Finally taking the $p-$ th root on both sides and using the Minkowski's inequality $[\|x + y\|_{L_p} \leq \|x\|_{L_p} + \|y\|_{L_p}]$, we get

$$\|S(a)\| \leq CK_1[\sqrt{p}\|a\|_2 + p\|a\|_\infty].$$

(d) **The $p\|a\|_\infty$ term is unavoidable.** Let X be a standard doubly exponential distribution supported on $\mathbb{R}$, i.e., $\mathbb{P}[X = x] = \frac{1}{2}e^{-|x|}$. Then $\|X\|_{L_p}^p = \int_{-\infty}^\infty |x|^p \frac{1}{2} e^{-|x|} dx = \int_0^\infty x^p e^{-x} dx = \Gamma(p+1) = p!$. For large p, using Striling's approximation we get $p! \approx \sqrt{2\pi p}(p/e)^p \approx (\sqrt{2\pi e})^p p^p$. Therefore, for large p, $\|X\|_{L_p} \approx cp$, for some constant c. Therefore for large p, $\|S(a)\|_{L_p} = \|X\|_{L_p} \approx cp$, hence for exponential random variables, which are not sub-Gaussian, the $\mathcal{O}(p)$ bound on the RHS can not be reduced to $\mathcal{O}(\sqrt{p})$.

5 Bounding quadratic forms (Hanson–Wright inequalities).

This exercise studies a canonical order-two U-statistic (a Gaussian"chaos") and shows how two matrix norms control its fluctuations. This is a special case of the Hanson–Wright inequality.

Let $g = (g_1, g_2, \dots, g_n) \sim \mathcal{N}(0, I_n)$ have independent standard normal coordinates, and let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix with zero diagonal ($a_{ii} = 0 \quad \forall \quad i$). Define the second-order Gaussian chaos

$$Z = g^\top A g = \sum_{i,j=1}^{n} a_{ij} g_i g_j = 2 \sum_{1\le i<j\le n} a_{ij} g_i g_j.$$

let $\|\cdot\|$ denotes the spectral/operator norm and $\|\cdot\|_F$ denotes the Frobenius norm:

$$\|A\| = \sup_{\|x\|_2=1} \|Ax\|_2; \quad \|A\|_F = \left(\sum_{i,j=1}^{n} a_{ij}^2\right)^{1/2}.$$

For symmetric A, $\|A\|$ is the largest absolute eigenvalue, and $\|A\|_F^2$ is the sum of squared eigenvalues.

(a) **Mean and Variance.** In the expression of Z, the terms g_i^2 are absent since $a_{ii} = 0$. Also g has independent coordinates, therefore using linearity and independence of expectation,

$$\mathbb{E}[Z] = 2 \sum_{1\le i<j\le n} a_{ij}\mathbb{E}[g_i]\mathbb{E}[g_j] = 0 \quad [\text{Since } \mathbb{E}[g_i] = 0 \forall i.]$$

we can also use that $\mathbb{E}[g^\top A g] = \mathbb{E}[\mathrm{tr}(g^\top A g)] = \mathbb{E}[\mathrm{tr}(gg^\top A)] = \mathrm{tr}[\mathbb{E}[gg^\top]A] = \mathrm{tr}[I_n A] = 0.$

Since $\mathbb{E}[Z] = 0$, $\mathrm{Var}[Z] = \mathbb{E}[Z^2]$. Hence,

$$\begin{aligned}
\mathrm{Var}[Z] = \mathbb{E}[Z^2] &= \mathbb{E}[\sum_{i,j} a_{ij} g_i g_j)(\sum_{k,l} a_{kl} g_k g_l)] \\
&= \mathbb{E}[\sum_{i,j,k,l} a_{ij} a_{kl} g_i g_j g_k g_l] \\
&= \sum_{i,j,k,l} a_{ij} a_{kl} \mathbb{E}[g_i g_j g_k g_l] \\
&= \sum_{i,j,k,l} a_{ij} a_{kl} [\delta_{ij}\delta_{kl} + \delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}] \\
&= \sum_{i,j} a_{ij} a_{ij} + \sum_{i,l} a_{il} a_{il} \\
&= 2\left(\sum_{i,j} a_{ij}^2\right) \\
&= 2\|A\|_F^2.
\end{aligned}$$

(b) **Variance Bound Using Gausian Poincaré.** For $f(x) = x^\top A x$, $\nabla f(x) = 2Ax$ (when A is symmetric), Using this we get, $\nabla f(g) = 2Ag$. Hence using the Poincaré inequality,

$$\mathrm{Var}[f(g)] \leq \mathbb{E}[\|\nabla f(g)\|_2^2] = \mathbb{E}[\|2Ag\|_2^2] = 4\mathbb{E}[(g^\top A^\top Ag)] = 4\mathbb{E}[\mathrm{tr}[A^\top Agg^\top]] = 4\,\mathrm{tr}[A^\top A\mathbb{E}(gg^\top)] = 4\,\mathrm{tr}[A^\top AI_n] = 4\|A\|_F^2.$$

In the proof of Convex Poincaré Inequality, we derived the variance bound using point(co-ordinate) wise range bound, here each g_i has a variance of 1, which couples with the multiplication with a_{ij}, this gives an additional $\pm a_{ij}$ factor, squaring it gives a factor of 4.

(c) **Diagonalization and Distributional Representation.** For an orthogonal matrix U, $Ug \overset{\mathrm{d}}{=} g \sim \mathcal{N}(0, I_n)$. This is a simple transformation of random variable. Therefore for $A = U^\top \Lambda U$, and using $\mathrm{tr}[A] = \sum_{i=1}^n \lambda_i = 0$

$$Z = g^\top Ag = (Ug)^\top \Lambda (Ug) \overset{\mathrm{d}}{=} g^\top \Lambda g = \sum_{i=1}^n \lambda_i g_i^2 = \sum_{i=1}^n \lambda_i g_i^2 - \sum_{i=1}^n \lambda_i = \sum_{i=1}^n \lambda_i (g_i^2 - 1).$$

(d) **CGF of a centered χ_1^2.** Let $G \sim \mathcal{N}(0,1)$. Hence, $G^2 \sim \chi_1^2$. We will use the facts: MGF of a $R \sim \mathrm{Gam}(\alpha, \beta)$ is $\mathbb{E}[e^{\theta R}] = (1-\theta\beta)^{-\alpha}$, for $\theta < 1/\beta$. Also, $\chi_1^2 = \mathrm{Gam}(1/2, 2)$. Therefore, the MGF of a $T \sim \chi_1^2$ is $\mathbb{E}[e^{\theta T}] = (1-2\theta)^{-1/2}$ for $\theta < 1/2$. Hence for $\theta < 1/2$,

$$\log \mathbb{E}[e^\theta Y] = \log \mathbb{E}[e^{\theta(G^2-1)}] = \log \mathbb{E}[e^{-\theta} e^{\theta G^2}] = -\theta + \log((1-2\theta)^{-1/2}) = -\theta - \frac{1}{2}\log(1-2\theta).$$

For $\theta > 0$, set $x = 2\theta \in [0,1)$. The series expansion of $-\log(1-x)$ at $x = 0$ is $\sum_{i=1}^\infty \frac{x^i}{i}$. Therefore, $-\log(1-x) - x = x^2/2 + x^3/3 + x^4/4 + ... < \frac{1}{2}\frac{x^2}{1-x} = \frac{2\theta^2}{1-2\theta}$. So,

$$\log \mathbb{E}[e^{\theta Y}] = -\theta - \frac{1}{2}\log(1-2\theta) = \frac{1}{2}[-2\theta - \log(1-2\theta)] \leq \frac{1}{2}\frac{2\theta^2}{1-2\theta} = \frac{\theta^2}{1-2\theta}.$$

Also, when $\theta < 0$, $(2\theta)_+ = 0$. In this case, we can use $-\log(1-x) - x < x^2/2 = 2\theta^2$. Hence,

$$\log \mathbb{E}[e^{\theta Y}] = \frac{1}{2}[-2\theta - \log(1-2\theta)] \leq \frac{1}{2}2\theta^2 = \theta^2 = \frac{\theta^2}{1-(2\theta)_+}.$$

Combining both cases, we got the required bound.

(e) **A Bernstein-type CGF Bound.** using part (c) we have $Z \overset{\mathrm{d}}{=} \sum_{i=1}^n \lambda_i(g_i^2 - 1)$. Now, $g_i \sim \mathcal{N}(0,1)$. So, using part (d), we can write for all i, $\log \mathbb{E}[e^{\lambda\lambda_i(g_i^2-1)}] \leq \frac{(\lambda\lambda_i)^2}{1-2\lambda\lambda_i}$ [this is satisfied by the λ bound, since $\lambda_i \leq \lambda_{\max} = \|A\|$, so, $\lambda\lambda_i < \frac{1}{2\|A\|}\lambda_i \leq \frac{1}{2\|A\|}\|A\| = \frac{1}{2}$]. Combining this with the factorization of the log-MGF, we get

$$\log \mathbb{E}[e^{\lambda Z}] = \log \mathbb{E}[e^{\sum_{i=1}^n \lambda\lambda_i(g_i^2-1)}] = \sum_{i=1}^n \log \mathbb{E}[e^{\lambda\lambda_i(g_i^2-1)}] \leq \sum_{i=1}^n \frac{(\lambda\lambda_i)^2}{1-2\lambda\lambda_i} \leq \sum_{i=1}^n \frac{(\lambda\lambda_i)^2}{1-2\lambda\|A\|} = \sum_{i=1}^n \frac{\lambda^2\|A\|_F^2}{1-2\lambda\|A\|},$$

Since, $\lambda_i \leq \lambda_{\max} = \|A\| \implies 1 - 2\lambda\lambda_i \geq 1 - 2\lambda\|A\| \implies \frac{1}{1-2\lambda\lambda_i} \leq \frac{1}{1-2\lambda\|A\|}$ and $\|A\|_F^2 = \mathrm{tr}(A^\top A) = \mathrm{tr}(\Lambda^\top \Lambda) = \sum_{i=1}^n \lambda_i^2$. This also holds for $-\frac{1}{2\|A\|} < \lambda < 0$. Hence, $\log \mathbb{E}[e^{\lambda Z}] \leq \frac{\lambda^2\|A\|_F^2}{1-2|\lambda|\|A\|}$.

(f) **Tail Bound.** Using Chernoff bound along with part (d), we get

$$\mathbb{P}[Z \geq t] \leq \inf_{\lambda>0} \exp(-\lambda t + \log \mathbb{E}e^{\lambda Z}) \leq \inf_{\lambda>0} \exp\left(-\lambda t + \frac{\lambda^2 \|A\|_F^2}{1-2|\lambda|\|A\|}\right) \leq \exp\left(-\lambda_0 t + \frac{\lambda_0^2 \|A\|_F^2}{1-2|\lambda_0|\|A\|}\right),$$

for some $\lambda_0 \in [0, 1/(2\|A\|))$, which we will choose so that the right hand side becomes $-u$ when $t = 2\|A\|_F\sqrt{u} + 2\|A\|u$. Note that for $\lambda_0 = \frac{\sqrt{u}}{\|A\|_F + 2\|A\|\sqrt{u}} < \frac{1}{2\|A\|}$,

$$\begin{aligned}
-\lambda_0 t + \frac{\lambda_0^2 \|A\|_F^2}{1-2|\lambda_0|\|A\|} &= -\frac{\sqrt{u}}{\|A\|_F + 2\|A\|\sqrt{u}} t + \frac{\frac{u}{(\|A\|_F+2\|A\|\sqrt{u})^2}\|A\|_F^2}{1-2\frac{\sqrt{u}}{\|A\|_F+2\|A\|\sqrt{u}}\|A\|} \\
&= -\frac{\sqrt{u}t}{\|A\|_F + 2\|A\|\sqrt{u}} + \frac{u\|A\|_F}{\|A\|_F + 2\|A\|\sqrt{u}} \\
&= -\frac{\sqrt{u}\sqrt{u}(2\|A\|_F + 2\|A\|\sqrt{u})}{\|A\|_F + 2\|A\|\sqrt{u}} + \frac{u\|A\|_F}{\|A\|_F + 2\|A\|\sqrt{u}} \\
&\leq -2u\frac{\|A\|_F + 2\|A\|\sqrt{u}}{\|A\|_F + 2\|A\|\sqrt{u}} + u \\
&\leq -u.
\end{aligned}$$

Hence, $\mathbb{P}[Z \geq 2\|A\|_F\sqrt{u} + 2\|A\|u] \leq \exp(-u)$.

For the Bernstein looking bound note that, for $t \leq 2\|A\|_F\sqrt{u} + 2\|A\|u$, it is true that $u \geq c\min\left\{\frac{t^2}{\|A\|_F^2}, \frac{t}{\|A\|}\right\}$. We also use the identity $\min(x,y) \geq \frac{xy}{x+y}$. Using that, we get

$$\begin{aligned}
\min\left\{\frac{t^2}{\|A\|_F^2}, \frac{t}{\|A\|}\right\} &\geq \frac{t^3/(\|A\|_F^2\|A\|)}{(\|A\|t^2 + \|A\|_F^2 t)/(\|A\|_F^2\|A\|)} \\
&= \frac{t^2}{\|A\|_F^2 + \|A\|t}.
\end{aligned}$$

Therefore,

$$\mathbb{P}[Z \geq t] \leq \exp\left(-c\frac{t^2}{\|A\|_F^2 + \|A\|t}\right) \lesssim \exp\left(\frac{-t^2/4}{\|A\|_F^2 + \|A\|t}\right).$$

(g) L_p **Bound.** We will use the fact that for any $a, b > 0$, $\exp(-c\min\{a,b\}) \leq e^{-a} + e^{-b}$. Hence $\mathbb{P}[|Z| \geq t] \leq 2\exp(-t^2/\|A\|_F^2) + 2\exp(-t/\|A\|)$. Using the same calculation for tail bound to moment growth implication in Problems 2 and 4, we know that $\exp(-t^2/\|A\|_F^2)$ gives rise to moment growth of order $\sqrt{p}$ (behaves like the moment growth for sub-Gaussian variables) and $\exp(-t/\|A\|)$ gives rise to moment growth of order p (behaves like sub-exponential). Combining them with the associated $\|Z\|_{\psi_1} \lesssim \|A\|$ (for the sub-exponential growth) and the $\|Z\|_{\psi_2} \lesssim \|A\|_F$ (for the sub-Gaussian growth), we get the resultant moment growth:

$$\|Z\|_{L_p} = \mathbb{E}[|Z|^p]^{1/p} \leq C(\|A\|_F\sqrt{p} + \|A\|p).$$